DOCUMENT RESUME

ED 231 881 TM 830 493

| | |
|---|---|
| AUTHOR | Mumford, Michael D.; Mendoza, Jorge L. |
| TITLE | Range Restriction and Attenuation Corrections. |
| SPONS AGENCY | National Institutes of Health (DHHS), Bethesda, Md. |
| PUB DATE | Mar 83 |
| GRANT | NIH-HD-14244-02 |
| NOTE | 15p.; Paper presented at the Annual Meeting of the Southeastern Psychological Association (Atlanta, GA, March 23-26, 1983). |
| PUB TYPE | Speeches/Conference Papers (150) -- Reports - Research/Technical (143) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Correlation; *Measurement Techniques; Psychometrics; *Research Problems; *Statistical Analysis; *Statistical Bias; Test Reliability; Test Theory |
| IDENTIFIERS | *Correction for Attenuation; Magnitude Estimations; *Range Restriction; Sequential Estimation |

ABSTRACT

The present paper reviews the techniques commonly used to correct an observed correlation coefficient for the simultaneous influence of attenuation and range restriction effects. It is noted that the procedure which is currently in use may be somewhat biased because it treats range restriction and attenuation as independent restrictive influences. Subsequently, an equation was derived which circumvents this difficulty and provides a more general solution to the problem of estimating the true magnitude of a correlation coefficient in data sets where these restrictive influences are operating. Finally, the nature of the bias induced by application of the common corrective technique is identified and related to the equation derived in the study at hand. (Author)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED231881

# Range Restriction And Attenuation Corrections

Michael D. Mumford And Jorge L. Mendoza

Advanced Research Resources Organization The University Of Georgia

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. D. Mumford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 498

Acknowledgements

This research has been supported in part by N.I.H. Grant H.D. 14244-02. Requests for reprints should be sent to Dr. Jorge L. Mendoza, Department of Psychology, The University of Georgia, Athens, Georgia 30602.

## Abstract

The present paper reviews the techniques commonly used to correct an observed correlation coefficient for the simultaneous influence of attenuation and range restriction effects. It is noted that the procedure which is currently in use may be somewhat biased because it treats range restriction and attenuation as independent restrictive influences. Subsequently, an equation was derived which circumvents this difficulty and provides a more general solution to the problem of estimating the true magnitude of a correlation coefficient in data sets where these restrictive influences are operating. Finally, the nature of the bias induced by application of the common corrective technique is identified and related to the equation derived in the study at hand.

One of the most pervasive methodological problems facing psychologists entails determination of the techniques which are to be used in assessing the nature and strength of the relationship between various measures. Of course, the correlation coefficient has provided the field with a viable statistical tool for solving this problem in the univariate case. Unfortunately, in some instances the appropriateness of correlational techniques may be limited by the operation of certain statistical biases in actual data bases. Thorndike (1949) has noted that two of these biases, termed range restriction and attenuation effects, can exert a powerful diminishing influence on the magnitude of observed correlation coefficients. Range restriction occurs when the variability in a sample is reduced on one or more measures relative to that observed in the target population, as a result of the operation of spurious influences such as prior selection on an unmeasured but positively correlated extraneous variable. The net effect of range restriction is a reduction in the expected magnitude of the observed correlation coefficient and an underestimate of the true relationship between the variables. Attenuation effects refer to the fact that an observed correlation coefficient will tend to underestimate the true magnitude of the relationship between two variables to the extent that these measures are not an accurate reflection of true variation, i.e., to the extent that they are unreliable. In some applied studies the operation of these biases may be acceptable. Yet when an investigation centers on determining the true strength of the relationship between two sets of measures the operation of these biases in the experimental data base constitutes a serious, often unavoidable, confound.

Psychometrics has long been cognizant of the implications of range restriction and attenuation effects with respect to the inferrences drawn by investigators concerning the magnitude of relationships. Consequently, a

variety of equations have been derived which permit the investigator to correct data based estimates of the magnitude of a correlation coefficient for the operation of these influences (Guilford, 1954; Stanley, 1971). However, these equations were designed to correct an observed coeffieient for the operation of a single biasing influence. When a researcher is concerned with determining the magnitude of a relationship irrespective of both those range restriction and attenuation effects which may be operating, the following procedure is generally utilized. Initially, the observed correlation coefficient is corrected for attenuation in the predictor and/or criterion measures via this formula

$$r_{cA} = \frac{r_{xy}}{\sqrt{r_{xx}} \cdot \sqrt{r_{yy}}},$$

where:

$r_{cA}$ = the correlation between the predictor and criterion corrected for attenuation

$r_{xy}$ = the observed correlation between the predictor and criterion

$r_{xx}$ = the reliability of the predictor

$r_{yy}$ = the reliability of the criterion.

Once $r_{cA}$ has been obtained, this term is entered into the particular equation correcting for the effects of range restriction which is appropriate with respect to the methodological situation at hand. In those cases where range restriction has occurred because of prior selection on the basis of predictor scores, Thorndike's Type II, the following equation would be used,

$$r_{cRA} = \frac{r_{cA} \cdot \frac{6x(R)}{6x}}{\sqrt{1 - r_{cA}^2 + r_{cA}^2 \cdot \left(\frac{6_x^2(R)}{6_x^2}\right)}},$$

where:

$r_{cRA}$ = the correlation between the predictor and criterion corrected for attenuation and range restriction

$6^2_{x(R)}$ = the variance of the predictor in the restricted sample

$6^2_x$ = the variance of the predictor in the unrestricted sample

$6_{x(R)}$ = the standard deviation of the predictor in the restricted sample

$6_x$ = the standard deviation of the predictor in the unrestricted sample

While the correction of an observed correlation coefficient for the effec of range restriction and attenuation through the use of this sequential strate has seen wide application in both theoretical and applied studies, the appropr ness of this procedure is open to question on the basis of at least three cons derations. First, application of the sequential strategy implicity assumes that range restriction and attenuation operate as independent biasing effects. However, as Magnusson (1966) has pointed out, because range restriction acts to reduce true variation while leaving error variance constant, it tends to deflate reliability as well as validity estimates. The implication here is that range restriction and attenuation represent correlated rather than inde- pendent biasing effects. This in turn suggests, that the sequential correctio strategy outlined above yields a biased estimate of the true magnitude of the relationship between a predictor and criterion. Second, implementations of th sequential strategy generally utilize only one of the special case corrections for range restriction, and since multiple types of restriction may operate in study, application of this strategy can result in some degree of underestimati of the true strength of the predictor, criterion relationship. It is of note that this observation indicates the need for a more general solution. Finally all corrections for range restriction assume that the slope of the best fittin regression line between the predictor and criterion measures is identical in both the restricted and unrestricted samples (Ghiselli, Campbell and Zedeck, 1981). To the extent range restriction operates to reduce true score variatic while the error variance remains constant, this assumption will only rarely be met since $b = r_{xy} \cdot \frac{6y}{6x}$. This suggests the presence of a further biasing influen

in the sequential strategy and current corrections for range restriction.

Any bias arising from the foregoing assumptional violation may be eliminated if it is assumed that the regression of predictor true scores on criterion true scores is constant within both the restricted and the unrestricted samples. In the present paper an attempt was made to utilize this assumption in order to to derive a general solution for correcting an observed correlation coefficient for the simultaneous operation of attenuation and multiple range restriction effects. Additionally, an attempt was made to demonstrate the nature of the bias which arises through application of the sequential correction strategy.

The assumption of a constant true score regression line between the predictor and criterion measures, regardless of the degree and kind of range restriction, implies that the correlation between true scores within the restricted and unrestricted samples may be determined through the following equations, under conditions of linearity, normality and homoscedacity of true scores;

$$(1A)\quad r^2_{txty} = b^2_{txty} \cdot \frac{6^2_{tx}}{6^2_{ty}} \qquad (1B)\quad r^2_{txty(R)} = b^2_{txty} \cdot \frac{6^2_{tx(R)}}{6^2_{ty(R)}}$$

where:

$b^2_{txty}$ = the square of the regression of true predictor scores on true criterion scores

$r^2_{txty}$ = the square of the correlation between predictor and criterion true scores in the unrestricted sample

$r^2_{txty(R)}$ = the square of the correlation between predictor and criterion true scores in the restricted sample

$6^2_{tx}$ = the variance of true scores on the predictor in the unrestricted sample

$6^2_{tx}$ = the variance of true scores on the criterion in the unrestricted sample

$6^2_{tx(R)}$ = the variance of true scores on the predictor in the restricted sample

$6^2_{tx(R)}$ = the variance of true scores on the criterion in the restricted sample

Simple algebraic transformation of equations (1A) and (1B) yields

$$(2A)\quad b^2_{txty} = r^2_{txty} \cdot \frac{6^2_{ty}}{6^2_{tx}} \qquad (2B)\quad b^2_{txty} = r^2_{txty(R)} \frac{6^2_{ty(R)}}{6^2_{tx(R)}}.$$

According to our initial assumption this implies that

$$(3)\quad r^2_{txty} \cdot \frac{6^2_{ty}}{6^2_{tx}} = r^2_{txty(R)} \cdot \frac{6^2_{ty(R)}}{6^2_{tx(R)};}$$

which in turn leads to the following expressions

$$(4A)\quad r^2_{txty(R)} = r^2_{txty} \cdot \frac{6^2_{ty}}{6^2_{tx}} \cdot \frac{6^2_{tx(R)}}{6^2_{ty(R)}}$$

$$(4B)\quad r^2_{txty} = r^2_{txty(R)} \cdot \frac{6^2_{tx}}{6^2_{ty}} \cdot \frac{6^2_{ty(R)}}{6^2_{tx(R).}}$$

Equations (4A) and (4B) specify the relationships between true score correlations obtained for two measures in a restricted and unrestricted sample, regardless of the degree and kind of range restriction. In the present investigation equation (4B) is of particular interest since it specifies the formula for simultaneously estimating the unattenuated, unrestricted correlation between a predictor and criterion measure on the basis of data obtained within the restricted sample and knowledge of the variance of these measures within the unrestricted sample. However, this equation has little practical value in the assessment of the true strength of the relationship between predictor and criterion scores since the right hand terms are expressed as a function of unobservable true scores. An initial step in eliminating this difficulty may be taken by rewriting the true score variances contained in equation (4B) in terms of the relevant observed variances and reliability coefficients. This substitution yields the following equation

$$(5)\quad r^2_{txty} = r^2_{txty(R)} \cdot \frac{r_{xx}6^2_x}{r_{yy}6^2_y} \cdot \frac{r_{yy(R)}\,6^2_{y(R)}}{r_{xx(R)}\,6^2_{x(R);}}$$

which when rearranged leads to

$$(6)\quad r^2_{txty} = r^2_{txty(R)} \cdot \frac{r_{xx}}{r_{yy}} \cdot \frac{r_{yy(R)}}{r_{xx(R)}} \cdot \frac{6^2_{y(R)}}{6^2_{y}} \cdot \frac{6^2_{x}}{6^2_{x(R)}},$$

where $r_{xx}$ and $r_{yy}$ are the reliabilities of the predictor and criterion measures in the unrestricted sample, and $r_{xx(R)}$ and $r_{yy(R)}$ reflect the reliabilities of these measures in the restricted sample. At this point there remains only the question as to how the term $r^2_{txty(R)}$ can be estimated from observable relationships. This issue can be resolved by noting that the correlation observed in a restricted sample is defined by the equation

$$(7)\quad r^2_{xy(R)} = \frac{b^2_{txty} \cdot 6^4_{tx(R)}}{6^2_{x(R)} \cdot 6^2_{y(R)}},$$

since $cov^2(x,y)R = cov\ (tx,ty)R = b^2_{txty(R)} 6^4_{tx(R)}$ and $b^2_{txty(R)} = b^2_{txty}$. Multiplying both sides of equation (7) by $\frac{6^2_{tx(R)}}{6^2_{tx(R)}} \cdot \frac{6^2_{ty(R)}}{6^2_{ty(R)}}$ and simplifying, one obtains

$$(8)\quad r^2_{txty(R)} = \frac{r^2_{xy(R)}}{r_{xx(R)} \cdot r_{yy(R)}}.$$

When the foregoing expression is substituted into equation (6) $r^2_{txty}$ may be rewritten as

$$(9)\quad r^2_{txty} = \frac{r^2_{xy(R)}}{r_{xx(R)} \cdot r_{yy(R)}} \cdot \left(\frac{r_{xx}}{r_{yy}}\right) \cdot \left(\frac{r_{yy(R)}}{r_{yy}}\right) \cdot \left(\frac{6^2_{y(R)}}{6^2_{y}}\right) \cdot \left(\frac{6^2_{x}}{6^2_{x(R)}}\right).$$

Equation (9) presents the formula for the simultaneous correction of attenuation and range restriction effects on the basis of observable information. Most of the statistical information required for implementation of this equation should be readily available to the investigator. The variance of scores on the predictor and criterion within the sample being examined in the research effort

may be used as estimates of the restricted variances of these measures. Estimates of the unrestricted variance of the predictor and criterion measures, as well as their unrestricted reliabilities, may be obtained from the standard sources of normative information. The only parameters required by this equation which might not be available are the estimates of the reliability of the predictor and criterion measures, because investigators commonly make no attempt to obtain this information. However, the lack of direct estimates for these two parameters does not necessarily preclude the use of equation (9) if it is assumed that the standard error of measurement, or the variance around a true score, is constant in both the restricted and unrestricted samples on the predictor and criterion measures. Under these conditions the restricted reliabilities may be estimated by the following formulas

$$(10) \quad r_{xx(R)} = 1 - \left( \frac{\sigma^2_x}{\sigma^2_{x(R)}} \cdot (1 - r_{xx}) \right)$$

$$(11) \quad r_{yy(R)} = 1 - \left( \frac{\sigma^2_y}{\sigma^2_{y(R)}} \cdot (1 - r_{yy}) \right) .$$

Thus the restricted reliabilities of the predictor and criterion measures can be estimated from data readily available to the investigator. As a result, it appears that there are no serious impediments to the use of equation (9) in correcting correlation coefficients for the operation of attenuation and range restriction effects.

Because it appears that it is possible to implement equation (9) in practice, it now seems appropriate to examine the relationship between this formula and the traditional sequential procedure. While the particular order in which the steps of the sequential are carried out is of little import, we will begin with the general correction for range restriction (e.g., see Gishelli, Campbell, and Zedeck, 1981); which specifies that

$$(12)\quad r^2_{cR} = r^2_{xy(R)} \cdot \frac{6^2_{y(R)}}{6^2_{x(R)}} \cdot \frac{6^2_x}{6^2_y} ,$$

where $r^2_{cR}$ denotes the square of the predictor criterion correlation coefficient corrected for range restriction. Multiplying both sides of equation (12) by $\frac{6^2_{tx}}{6^2_{ty}} \cdot \frac{6^2_{ty}}{6^2_{tx}}$ and employing equation (7) the following expressions are obtained

$$(13A)\quad r^2_{cR} = \frac{b^2_{txty} \cdot 6^4_{tx} \cdot 6^2_x}{6^4_{x(R)} \cdot 6^2_y} \cdot \frac{6^2_{tx}}{6^2_{ty}} \cdot \frac{6^2_{ty}}{6^2_{tx}}$$

$$(13B)\quad r^2_{cR} = \left(b^2_{txty} \cdot \frac{6^2_{tx}}{6^2_{ty}}\right) \cdot \left(\frac{6^2_{ty} \cdot 6^2_{tx(R)} \cdot 6^2_x}{6^2_{tx} \cdot 6^4_{x(R)} \cdot 6^2_y}\right) .$$

Since $b^2_{txty} \cdot \frac{6^2_{tx}}{6^2_{ty}} = r^2_{txty}$, substitution and rearrangement results in the expression

$$(14)\quad r^2_{cR} = r^2_{txty} \cdot \left(\frac{r_{yy}}{r_{xx}}\right) \cdot r^2_{xx(R)} .$$

Equation (14) presents the correlation coefficient corrected for range restriction. Now correcting both sides of equation (14) for attenuation yields

$$(15)\quad r^2_{cR} \cdot \frac{1}{r_{xx(R)} \cdot r_{yy(R)}} = r^2_{txty} \cdot \left(\frac{r_{yy}}{r_{xx}}\right) \cdot r^2_{xx(R)} \cdot \frac{1}{r_{xx(R)}\, r_{yy(R)}}$$

If $r^2_{cRA}$ is used to designate the square of the fully corrected sequential coefficient, then equation (15) may be simplified to the following expressions

$$(16A)\quad r^2_{cRA} = r^2_{txty} \cdot \left(\frac{r_{yy}}{r_{xx}}\right) \cdot \frac{r_{xx(R)}}{r_{yy(R)}}$$

$$(16B)\quad r^2_{txty} = r^2_{cRA} \cdot \left(\frac{r_{xx}}{r_{yy}}\right) \cdot \frac{r_{yy(R)}}{r_{xx(R)}} .$$

Earlier it was noted that our central concern in carrying out corrections for range restriction and attenuation is the reproduction of the true magnitude of the relationship between predictor and criterion scores after the effects of range restriction and attenuation have been removed. However, as equation (16B) demonstrates the sequential correction procedure will produce an estimate of $r^2_{txty}$ which differs systematically from $r^2_{txty}$ by a factor of $\frac{r_{xx}}{r_{yy}} \cdot \frac{r_{yy(R)}}{r_{xx(R)}}$. This implies that the sequential correction procedure will yield a biased estimate of $r^2_{txty}$ except in those rare cases where $\frac{r_{xx}}{r_{yy}} \cdot \frac{r_{yy(R)}}{r_{xx(R)}}$ is equal to one; that is, when there are no range restriction effects or the predictor and criterion measures are of equal reliability in both the restricted and unrestricted samples. Whether $r^2_{txty}$ is overestimated or underestimated will depend on the particular combination of predictor and criterion reliabilities obtained in the restricted and unrestricted samples. Yet it is of note that this bias can be substantial. For instance, if the unrestricted reliability of the predictor is .85, the unrestricted reliability of the criterion is .70 and there is a 33% restriction of range soley on the criterion measure, then application of the sequential correction strategy will yield a 22% overestimate of $r^2_{txty}$. Since the foregoing example is a reasonably realistic presentation of the conditions observed in the selection situation, it seems clear that application of the sequential strategy can yield an estimate of $r^2_{txty}$ which is sufficiently biased to be a cause for concern. Of course, equation (16B) may be used to remove this bias. However, since $r^2_{cRA}$ is equal to $\frac{r^2_{xy(R)}}{r_{xx(R)} \cdot r_{yy(R)}} \cdot \frac{6^2_{y(R)}}{6^2_y} \cdot \frac{6^2_x}{6^2_{x(R)}}$, this correction will yield the equation

$$r^2_{txty} = \frac{r_{xy(R)}}{r_{xx(R)} \cdot r_{yy(R)}} \cdot \frac{6^2_{y(R)}}{6^2_y} \cdot \frac{6^2_x}{6^2_{x(R)}} \cdot \frac{r_{xx}}{r_{yy}} \cdot \frac{r_{yy(R)}}{r_{xx(R)}} ;$$

which is equivalent to equation (9) or the formula for the simultaneous correction of range restriction and attenuation effects.

Given the evidence presented in the foregoing paragraphs for the biased nature of the estimates of $r^2_{txty}$ obtained from the sequential correction procedure, it would seem that this technique should be replaced with the formula for the simultaneous correction of range restriction and attenuation effects. The simultaneous procedure should produce a sounder estimate of $r^2_{txty}$. Additionally, the simultaneous procedure appears to offer a somewhat more general solution to the problem of estimating the true magnitude of the relationship between two measures. This technique is capable of incorporating multiple specific range restriction effects and its derivation does not assume truncation. Moreover, this equation can be applied regardless of the particular degree of attenuation and/or range restriction operating on the predictor and/or criterion measures. For example, when there is no restriction of scores on the predictor and criterion measures, the simultaneous equation will be reduced to the traditional correction for attenuation; that is, $\frac{r^2_{xy(R)}}{r_{xx(R)} \cdot r_{yy(R)}}$.

Finally, it should be noted that the nature of the simultaneous equation suggests that any attempt to correct for both range restriction and attenuation effects, when estimating the true magnitude of a relationship, must incorporate the fact that range restriction and attenuation are interactive biasing influences.

## References

Ghiselli, E. E., Campbell, J. P., and Zedeck, S. Measurement Theory for the Behavioral Sciences. San Francisco, California; W. H. Freeman and Co., 1981.

Guilford, J. P. Psychometric Methods. New York; McGraw-Hill Inc., 1954.

Magnusson, D. Test Theory. Reading, Massachusetts; Addison-Weseley Inc., 1966.

Stanley, J. C. Reliability. In R. L. Thorndike (Ed.), Educational Measurement. Washington D. C.; American Council on Education, 1971.

Thorndike, R. L. Personnel Selection: Test and Measurement Techniques. New York; Wiley, 1949.